



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica
y del Medio Natural

Caracterización de las diferencias de sexo en esclerosis
múltiple, mediante la revisión sistemática y metaanálisis de
estudios transcriptómicos a nivel de microRNAs

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Muñoz del Campo, Ester

Tutor/a: Vilanova Navarro, Santiago

Cotutor/a externo: GARCIA GARCIA, FRANCISCO

Director/a Experimental: CATALA SENENT, JOSE FRANCISCO

CURSO ACADÉMICO: 2021/2022

Characterising sex differences in multiple sclerosis by systematic review and meta-analysis of microRNA transcriptomic studies

Abstract

Multiple sclerosis is a disease which implies central nervous system demyelination and neurodegeneration. Its effects differ regarding beginning, progress and clinical features; moreover, specific differences are spotted between male and female patients. Some of these sex-related differences include a higher risk to develop the disease in women, and a faster and more severe progression in men.

Seeking a better understanding of these sex-related disease differences, a systematic review and meta-analysis of microRNAs studies focused in multiple sclerosis (MS) was carried on. The systematic review allowed the identification of 6 studies which included the information needed to perform the analysis, however, after data evaluation only 4 of them were eligible. These studies' integrative analysis pointed out 4 microRNAs which presented an opposite expression profile between sexes, and also 17 microRNAs with the same profile in men and women were identified, representing a set of sex-independent MS markers. Functional characterization of these results provided sex-specific information about the biological functions of microRNAs' target genes.

This study has provided a better understanding of the disease's molecular mechanisms through marker identification, as well as the generation of new approaches for diagnosis and treatment, enhancing the development of personalized medicine.

Keywords: Multiple Sclerosis; MS; Sex differences; microRNA; Gene expression; Transcriptomics; Meta-Analysis; Personalized medicine

Author: Ester Muñoz del Campo

Academic tutor: Santiago Vilanova Navarro

External tutors: Francisco García García and José Francisco Català Senent

València, June 2022



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.

Caracterización de las diferencias de sexo en esclerosis múltiple, mediante la revisión sistemática y metaanálisis de estudios transcriptómicos a nivel de microRNAs

Resumen

La esclerosis múltiple es una enfermedad que conlleva la desmielinización y neurodegeneración del sistema nervioso central. Sus síntomas presentan variaciones en cuanto al inicio, progreso y manifestaciones clínicas entre pacientes; además hay determinados aspectos que difieren entre hombres y mujeres. En concreto, las mujeres presentan mayor riesgo de desarrollar la enfermedad, mientras que en los hombres la progresión es más rápida y severa.

Con el fin de estudiar y comprender mejor estas diferencias asociadas al sexo, se llevó a cabo un estudio que incluyó una revisión sistemática de datos de microRNAs procedentes de estudios centrados en esclerosis múltiple, y un posterior metaanálisis de estos. Mediante la revisión sistemática se identificaron 6 estudios con la información necesaria para llevar a cabo el análisis; aunque finalmente, tras la evaluación de los datos, sólo fueron elegibles 4 de ellos. El análisis integrado de estos estudios identificó 4 microRNAs con perfiles de expresión opuestos entre sexos, además, también se identificaron 17 microRNAs que presentan el mismo perfil en hombres y mujeres, por lo que constituyen un conjunto de marcadores de esclerosis múltiple independientes del sexo. La caracterización funcional de estos resultados proporcionó información específica por sexo, sobre las funciones biológicas en los genes diana de estos microRNAs.

En definitiva, la identificación de marcadores de la enfermedad en este trabajo ha proporcionado un mejor conocimiento de los mecanismos moleculares de esta patología, así como la generación de nuevas estrategias para su detección y tratamiento, potenciando el desarrollo de la medicina personalizada.

Palabras clave: Esclerosis múltiple; MS; Diferencias de sexo; microRNA; Expresión génica; Transcriptómica; Metaanálisis; Medicina personalizada

Autor: Ester Muñoz del Campo

Tutor académico: Santiago Vilanova Navarro

Cotutores externos: Francisco García García y José Francisco Català Senent

València, junio de 2022



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.

Agradecimientos

En primer lugar, agradezco mucho a Santi Vilanova y a su grupo del COMAV por acceder a guiarme y empezar a formarme como científica.

También me gustaría agradecer a Paco García y José Català por acogerme, ayudarme y enseñarme tanto durante el desarrollo del trabajo; y por darme la oportunidad de acercarme a un grupo científico de gran calidad tanto en proyectos como en personas.

Por último, gracias a todas esas personas que me han ayudado a progresar y a sacar adelante todos los retos que se nos presentaban. Gracias a mis entrenadores Cuca y José, por apostar por mí en los momentos en los que ni siquiera yo lo hago. Gracias a Alba, Clara y Maribel, por haber sido mi casa lejos de casa. Gracias a Edurne y Rocío, por convertirse en mis referentes más cercanas. Gracias a Patricia, María, Marta y Victoria, por adoptarme este último año. Gracias a mi hermana Lidón y mis tíos Esther y José, por inspirarme siempre a ser un poquito mejor. Infinitas gracias a mi madre Mayte, por hacer lo imposible con tal de darme la oportunidad de aprender.

Índice general

1. Introducción	10
1.1 Esclerosis múltiple.....	10
1.1.1. Subtipos de esclerosis múltiple.....	11
1.1.2. Epidemiología y componente de sexo.....	11
1.1.3. Factores de riesgo.....	12
1.1.4. Herramientas de diagnóstico.....	12
1.1.5. Tratamientos actuales.....	12
1.2 Datos ómicos y tecnologías de alto rendimiento.....	13
1.2.1. Transcriptómica.....	15
1.2.2. Análisis bioinformáticos de datos de expresión.....	15
1.2.3. Metaanálisis.....	15
2. Objetivos	17
3. Materiales y Métodos	18
3.1. Revisión sistemática.....	18
3.2. Análisis individual de los estudios.....	20
3.3. Metaanálisis.....	21
4. Resultados	23
4.1. Revisión sistemática.....	23
4.2 Análisis individual de los estudios.....	24
4.3 Metaanálisis.....	25
5. Discusión	29
6. Conclusiones	32
7. Bibliografía	33
8. Anexo de figuras	36

Índice de tablas

Tabla 4.1 Selección final de estudios para el metaanálisis. Se incluyen identificador, base de datos de procedencia, plataforma de estudio transcriptómico y referencia.

Tabla 4.2 Recuento de individuos pertenecientes a los grupos experimentales en los estudios seleccionados para el metaanálisis.

Tabla 4.3 Número de microRNAs significativos con sobreexpresión (↑) o expresión inhibida (↓) en cada estudio.

Tabla 4.4 Resumen de los microRNAs resultantes de la intersección entre los perfiles de expresión diferencial (↑: sobreexpresión; ↓: inhibición).

Tabla 4.5 Resumen de las rutas KEGG y términos GO con el enriquecimiento más significativo de cada perfil de expresión (↑: sobreexpresión; ↓: inhibición).

Índice de figuras

Figura 1.1 La sucesión de episodios de brote conducen a la acumulación de discapacidad en un período de 10 a 15 años. Se presenta la discapacidad, la actividad diagnosticada por imágenes de resonancia magnética (MRI) y los brotes (Doshi & Chataway, 2016).

Figura 1.2 Dogma Central de la Biología Molecular (elaboración propia).

Figura 1.3 Jerarquía de la evidencia (adaptado de Haidich, 2010).

Figura 3.1 Diagrama de las diferentes fases del trabajo, revisión sistemática, análisis individual y metaanálisis con sus respectivas etapas (♀: mujeres; ♂: hombres).

Figura 3.2 Diagrama de flujo de revisión sistemática según los estándares PRISMA (adaptado de Page et al., 2021).

Figura 3.3 Descripción de las cuatro comparaciones entre los estudios correspondientes al primer análisis (↑ para microRNAs sobreexpresados y ↓ para inhibidos, los datos comparados se rodean con líneas discontinuas).

Figura 3.4 Descripción de la comparación realizada en el segundo análisis, entre cada perfil de expresión sumando los datos aportados por cada estudio (↑ para microRNAs sobreexpresados y ↓ para inhibidos, los datos comparados se rodean con líneas discontinuas).

Figura 4.1 Resultados de la revisión sistemática de estudios de expresión en esclerosis múltiple, mediante la aplicación de los estándares de PRISMA.

Figura 4.2 Diagramas de Venn presentando las intersecciones entre los estudios GSE86863, GSE74579, GSE39643 y GSE21079 en microRNAs sobreexpresados en mujeres con MS (A), microRNAs sobreexpresados en hombres con MS (B), microRNAs inhibidos en mujeres con MS (C) y microRNAs inhibidos en hombres con MS.

Figura 4.3 Diagrama de Venn que compara los microRNAs con los distintos perfiles de expresión diferencial, señalando las intersecciones entre ellos. Se identifican las intersecciones que señalan los microRNAs con perfiles opuestos y con perfiles iguales en hombres y mujeres.

Figura 4.4 Selección de heatmaps en representación de los resultados del enriquecimiento funcional. Los microRNAs con perfil de expresión bajo en mujeres y alto en hombres presentan enriquecimientos con menor potencia estadística, tanto en rutas KEGG (A) como en términos GO (B); mientras que los demás, como los de perfil de expresión alto en mujeres y bajo en hombres (C) presentan enriquecimientos más claros.

Figura A.1 Heatmap de las rutas KEGG enriquecidas en aquellos microRNAs sobreexpresados en mujeres e inhibidos en hombres.

Figura A.2 Heatmap de las rutas KEGG enriquecidas en aquellos microRNAs inhibidos en mujeres y sobreexpresados en hombres.

Figura A.3 Heatmap de las rutas KEGG enriquecidas en los microRNAs sobreexpresados en pacientes de ambos sexos.

Figura A.4 Heatmap de las rutas KEGG enriquecidas en los microRNAs con baja expresión en pacientes de ambos sexos.

Figura A.5 Heatmap de los términos GO enriquecidos en aquellos microRNAs sobreexpresados en mujeres e inhibidos en hombres.

Figura A.6 Heatmap de los términos GO enriquecidos en aquellos microRNAs inhibidos en mujeres y sobreexpresados en hombres.

Figura A.7 Heatmap de los términos GO enriquecidos en los microRNAs sobreexpresados en pacientes de ambos sexos.

Figura A.8 Heatmap de los términos GO enriquecidos en los microRNAs sobreexpresados en pacientes de ambos sexos.

Abreviaturas

A	Adenina
C	Citosina
CIS	Síndrome clínico aislado (Clinically Isolated Syndrome)
DNA	Ácido desoxirribonucleico (Deoxyribonucleic Acid)
EBV	Virus de Epstein-Barr
ECM	Matriz extracelular (Extracellular Matrix)
GEO	Gene Expression Omnibus
GO	Gene Ontology
G	Guanina
KEGG	Kyoto Encyclopedia of Genes and Genomes
MRI	Imágenes de resonancia magnética (Magnetic Resonance Images)
mRNA	RNA mensajero
MS	Esclerosis múltiple (Multiple Sclerosis)
PB	Proceso biológico
PPMS	Esclerosis múltiple progresiva primaria (Primary Progressive Multiple Sclerosis)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RNA	Ácido ribonucleico (Ribonucleic Acid)
RNA-seq	Secuenciación de RNA
RRMS	Esclerosis múltiple con brotes y recuperación (Relapse-Remitting Multiple Sclerosis)
SNC	Sistema Nervioso Central
SPMS	Esclerosis múltiple progresiva secundaria (Secondary Progressive Multiple Sclerosis)
T	Timina
TF	Factor de transcripción (Transcription Factor)
U	Uracilo

1. Introducción

1.1. Esclerosis múltiple

La esclerosis múltiple (MS) es una enfermedad neuroinflamatoria con componente autoinmune que ataca a la mielina del sistema nervioso central (SNC) u oligodendrocitos. Se caracteriza por presentar episodios de brote y posterior recuperación con síntomas muy heterogéneos, incluyendo fatiga, espasmos, pérdida de visión, deterioro motor, dolor, alteraciones cognitivas y vértigo. Manifestándose entre los 20 y los 50 años con un pico a los 30, es una de las principales causas de discapacidad entre adultos jóvenes (Doshi & Chataway, 2016; Milo & Kahana, 2010). Mientras que el inicio y los brotes propios de la MS se achacan al componente autoinmune de la enfermedad, la acumulación progresiva de la discapacidad se asocia a los fenómenos neurodegenerativos. En la figura 1.1 se representa la acumulación de discapacidad a medida que se suceden los brotes de la enfermedad.

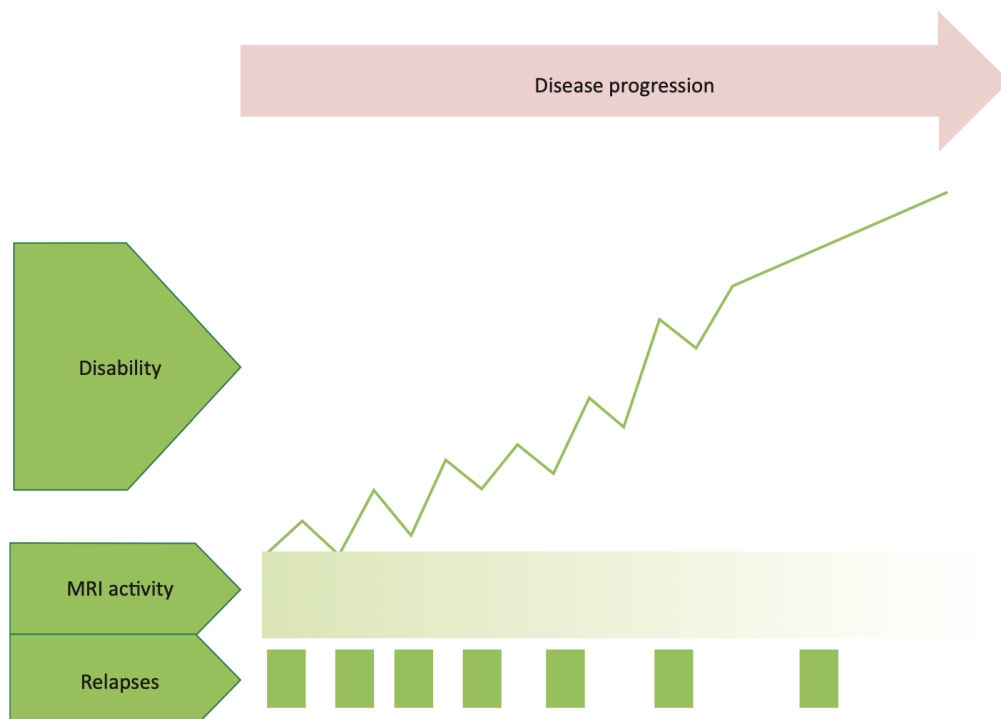


Figura 1.1. La sucesión de episodios de brote conducen a la acumulación de discapacidad en un período de 10 a 15 años. Se presenta la discapacidad, la actividad diagnosticada por imágenes de resonancia magnética (MRI) y los brotes (Doshi & Chataway, 2016).

Si bien la etiología de la MS no está claramente determinada, se sabe que en el desarrollo de la enfermedad intervienen factores genéticos, epigenéticos y ambientales (Thompson *et al.*, 2018). Al amplio espectro de síntomas y severidad se le suman variaciones en función del sexo del paciente. Existen diferencias en la prevalencia, siendo ésta mayor en las mujeres, aunque la progresión es más severa y veloz en el caso de los hombres. Otros aspectos distintivos son la edad en la que surge y los subtipos de la enfermedad que el paciente es más propenso a padecer en función del sexo (Voskuhl, 2020).

1.1.1. Subtipos de esclerosis múltiple

Existen distintos fenotipos que constituyen diferentes clases de MS, de las cuales las más comunes son: síndrome clínico aislado (clinically isolated syndrome, CIS), MS con brotes y recuperación (relapsing remitting MS, RRMS), MS progresiva primaria (PPMS) y MS progresiva secundaria (SPMS). La aparición de estos subtipos en un paciente se da de forma progresiva, es decir, por lo general se van sucediendo en el tiempo (Doshi & Chataway, 2016).

El CIS se corresponde con la manifestación inicial de la enfermedad, se da en forma de un ataque agudo en uno o varios lugares del SNC. En ocasiones los pacientes que han presentado CIS progresan a RRMS, que se caracteriza por una sucesión de brotes entre los que suele haber una recuperación. Entre 10 y 15 años después del diagnóstico de RRMS y la acumulación de recuperaciones incompletas, es común la transición a SPMS; la cual conlleva discapacidad y daño en la materia gris y blanca del sistema nervioso. Por último, PPMS afecta a entre el 10 y 15% de los pacientes, e implica una acumulación progresiva de discapacidad neurológica desde el inicio, normalmente debido a daños en la médula espinal (Doshi & Chataway, 2016; Milo & Kahana, 2010).

1.1.2. Epidemiología y componente de sexo

La prevalencia de la MS se encuentra entre 50 y 300 por cada 100.000 habitantes, aunque su distribución en el mundo no es regular. Se observa una tendencia a aumentar a medida que la zona geográfica se aleja del ecuador, aunque se encuentran excepciones (Thompson *et al.*, 2018).

También existen diferencias en función del sexo del paciente. Estas se observan en la susceptibilidad a padecer la enfermedad, que es de 2 a 3 veces mayor (dependiendo de la zona geográfica) por parte de las mujeres, quienes también presentan un inicio de la enfermedad en edades más tempranas. Alternativamente, la progresión de la enfermedad es más veloz y severa en hombres, ya que el tiempo transcurrido padeciendo los distintos subtipos progresivos de esclerosis múltiple es más corto (Voskuhl, 2020). En cuanto a los subtipos de MS, los hombres son más propensos a padecer PPMS, y por tanto, discapacidad desde el inicio (Thompson *et al.*, 2018).

No obstante, la variable biológica del sexo es habitualmente omitida en los diseños experimentales, de forma que se puede perder información clave del individuo en los estudios de enfermedades, afectando a la eficiencia del diagnóstico y del tratamiento. La MS es una de las condiciones donde el factor del sexo tiene unas implicaciones importantes y que deben ser estudiadas en mayor profundidad (Clayton & Collins, 2014). Por lo general, las respuestas inmunes de las mujeres son más fuertes que las de los hombres, hecho que se asocia al rol parental que estas desarrollan y que repercute en la susceptibilidad a padecer enfermedades con componente autoinmune, haciendo que su incidencia sea mayor en mujeres (Moulton, 2018). En concreto, los estudios como el de Voskuhl *et al.* (2018) señalan que las diferencias de sexo en MS pueden estar basadas en la expresión génica de los cromosomas sexuales en el SNC o en el sistema inmune, que a su vez puede estar modificada por las hormonas sexuales (Voskuhl, 2020; Voskuhl *et al.*, 2018).

1.1.3. Factores de riesgo

Los principales factores de riesgo de la MS son genéticos, epigenéticos y ambientales. Esto se deduce por la irregularidad en la distribución geográfica, y también por las variaciones de la incidencia en personas que migran a zonas geográficas con incidencia distinta a su lugar de origen. Es decir, que las personas procedentes de una zona geográfica con baja incidencia de MS presentan un aumento en la incidencia tras una generación al migrar a zonas donde la incidencia es mayor, y viceversa (Milo & Kahana, 2010).

En cuanto a los factores genéticos, se observa heredabilidad de la enfermedad, de forma que el riesgo aumenta cuanto más próximo sea el familiar afectado. Por otro lado, algunos factores ambientales que favorecen el desarrollo de la enfermedad son la deficiencia de vitamina D (que actúa como antiinflamatorio), la exposición a agentes infecciosos en edades tempranas (que disminuye las alergias y las respuestas autoinmunes) y el tabaquismo (aumenta el riesgo) (Thompson *et al.*, 2018; Milo & Kahana, 2010). También se ha asociado la infección por el virus de Epstein-Barr (EBV) tanto a la causa como a la progresión de la enfermedad; de hecho, recientemente se han publicado varios estudios como el de Bjornevik *et al.* (2022) que aportan grandes evidencias sobre esta asociación. La hipótesis se basa en la similitud de las proteínas del EBV con las del propio SNC, lo que desencadenaría la autoinmunidad de la MS (Bjornevik *et al.*, 2022; Houen *et al.*, 2020; Hrastelj & Robertson, 2022).

Todo ello junto con la diversidad de efectos patológicos, moleculares y clínicos hacen que la MS se considere una enfermedad multifactorial, lo que complica y a la vez hace muy necesario el estudio de los mecanismos biológicos subyacentes (Milo & Kahana, 2010).

1.1.4. Herramientas de diagnóstico

El diagnóstico de la MS se lleva a cabo combinando estrategias clínicas, de laboratorio y de obtención de imágenes. A nivel clínico es complicado diferenciar la MS de otras enfermedades neurológicas, por lo que es importante apoyarse también en las imágenes obtenidas por resonancia magnética (MRI) junto con tests serológicos (Thompson *et al.*, 2018). No obstante, aunque es cierto que la MRI puede ayudar a identificar la transición de CIS a RRMS, es una herramienta poco precisa debido a la heterogeneidad de sus dianas. Por ello, es útil para investigar la enfermedad pero para su uso en diagnóstico y monitorización del tratamiento presenta ciertas limitaciones, que se deberán superar debido a la naturaleza de su aplicación (Klineova & Lublin, 2018).

Las estrategias en desarrollo buscan complementar la información que ofrece el MRI con biomarcadores que permitan distinguir el subtipo y la progresión de la MS. En concreto, se ha comprobado que los microRNAs pueden realizar este papel, ya que permiten distinguir entre PPMS y RRMS. La identificación de nuevos biomarcadores se centra actualmente en las MS de tipo progresivo y en la predicción de la respuesta a los tratamientos, ya que son las necesidades más urgentes a nivel clínico (Thompson *et al.*, 2018).

1.1.5. Tratamientos actuales

Las estrategias actuales para tratar la MS se pueden clasificar en tres categorías: control del episodio agudo, tratamientos modificadores de la enfermedad y tratamientos de síntomas específicos (Doshi & Chataway, 2016).

Para manejar el episodio agudo se debe diferenciar si los síntomas se deben realmente a un episodio agudo o son causados por la desmielinización ya existente. Una vez confirmado el episodio, si presenta una severidad moderada o mayor, se administra metilprednisolona (corticoide) para acortar el episodio. Además, se hace una transfusión de sangre si la severidad es muy alta. Posteriormente se pueden emplear técnicas de fisioterapia para favorecer la recuperación (Doshi & Chataway, 2016).

En cuanto a los tratamientos modificadores de la enfermedad, los primeros surgieron en la década de los 90. Son moléculas como interferones-beta y anticuerpos humanizados. Todos ellos presentan efectos secundarios, y la decisión del tratamiento a aplicar se basa en si se prefiere empezar con un tratamiento más suave y progresar a medida que lo hace la enfermedad (estrategia de escalado), o si se elige empezar por un tratamiento más fuerte para evitar al máximo los brotes (estrategia de inducción) (Doshi & Chataway, 2016; Thompson *et al.*, 2018). La efectividad de estas estrategias se ve condicionada por el tipo de progresión de la enfermedad que presenta el paciente. Además, la elección del tratamiento modificador de la enfermedad adecuado es clave, ya que los tratamientos más eficaces son a la vez los que presentan efectos secundarios más agresivos (Thompson *et al.*, 2018).

La tercera categoría de tratamientos se centra en paliar los síntomas. Pueden ser de naturaleza farmacéutica o no. Por ejemplo se incluye hacer ejercicio y tener suficientes horas de sueño para la fatiga, o medicamentos como baclofen y gabapentin para los espasmos. No obstante, no todos los síntomas se pueden tratar de forma eficaz (Doshi & Chataway, 2016).

Estudios recientes como el de Houen *et al.*, 2020 señalan el papel que puede desarrollar el virus de Epstein-Barr (EBV) tanto en el desarrollo como en el progreso de la enfermedad. Consecuentemente se sugiere que un tratamiento antiviral podría influir beneficiosamente en el desarrollo y el curso clínico de la MS; por lo que se puede considerar un objetivo en el desarrollo de terapias (Bjornevik *et al.*, 2022).

Todo ello nos da una visión de las dificultades que surgen a la hora de lidiar con la MS, por lo que es conveniente alcanzar una mayor comprensión de la enfermedad. Esto permitiría desarrollar tratamientos y herramientas de diagnóstico lo más específicas y eficaces posible, avanzando hacia la medicina personalizada. Los objetivos más inmediatos incluyen frenar el avance de la enfermedad, encontrar tratamientos efectivos para la MS progresiva y también desarrollar estrategias para paliar los síntomas con el fin de mejorar el bienestar de los pacientes (Thompson *et al.*, 2018).

1.2. Datos ómicos y tecnologías de alto rendimiento

Las células almacenan la información heredable en forma de DNA (ácido desoxirribonucleico). Esta molécula es un heteropolímero bicatenario compuesto por combinaciones de cuatro monómeros. Cada uno de estos monómeros presenta una desoxirribosa, una unidad de fosfato y una base nitrogenada en la que difieren: adenina (A), timina (T), guanina (G) y citosina (C). La complementariedad de las dos cadenas de DNA establecida por los enlaces de tipo puente de hidrógeno entre las parejas de bases A-T y G-C, permiten que la información genética se transmita generacionalmente y también se exprese. Esto se debe a que tras la separación de las cadenas es posible emplear una de

ellas como molde, para sintetizar DNA o RNA (ácido ribonucleico). El RNA difiere respecto al DNA en la naturaleza del azúcar, en este caso ribosa, y en la sustitución de la base T por uracilo (U) (Kellenberger, 2004; Nelson & Cox, 2017).

Los diferentes tipos de RNA presentan diversas funciones, como la traducción a proteínas en el caso del RNA mensajero (mRNA), y también otras funciones reguladoras, como es el caso de los microRNAs, los cuales regulan la expresión génica. Este flujo de la información heredable de las células queda recogido en el Dogma Central de la Biología Molecular, que se presenta en la figura 1.2 (Nelson & Cox, 2017).

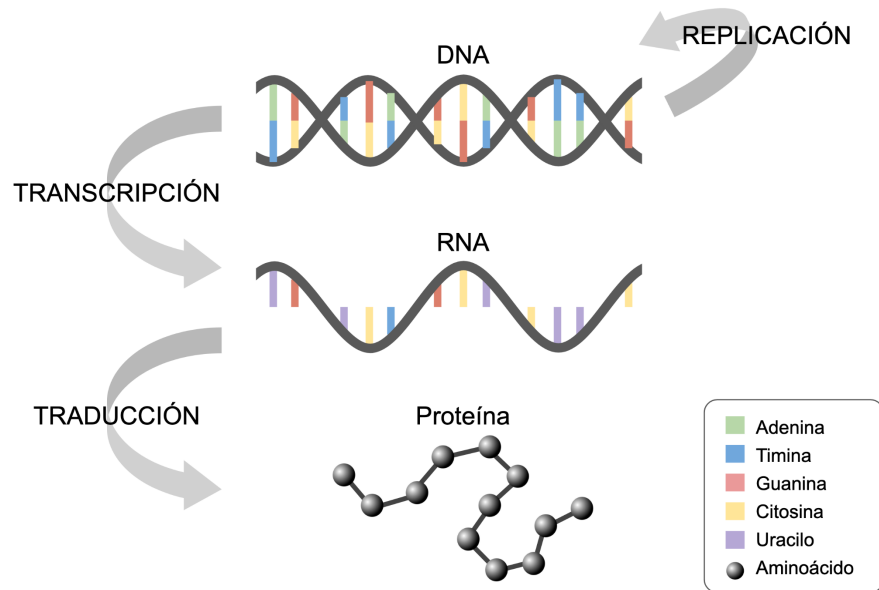


Figura 1.2. Dogma Central de la Biología Molecular (elaboración propia).

Por tanto, los sistemas biológicos se basan en la transferencia de información desde los ácidos nucleicos hasta las proteínas para poder funcionar. El estudio global de cada uno de los tipos de moléculas propios de los seres vivos es conocido como ciencias ómicas (Santiago-Rodríguez & Hollister, 2021). El estudio desde una perspectiva ómica ha sido posible gracias a los avances en tecnologías de alto rendimiento y la reducción de sus costes. Entre ellas se encuentran los arrays de expresión, basados en la hibridación de ácidos nucleicos a oligonucleótidos anclados a un chip, y también las herramientas de secuenciación de alto rendimiento como Next Generation Sequencing (Santiago-Rodríguez & Hollister, 2021; Hasin *et al.*, 2017).

Las ciencias ómicas incluyen genómica, transcriptómica, proteómica y metabolómica. Cada una de ellas se centra en un estamento de los organismos vivos, de forma que la genómica estudia el conjunto de la información del genoma, la transcriptómica estudia todo aquello que se plasma en RNA a partir del genoma, la proteómica estudia el conjunto de proteínas traducidas y la metabolómica estudia los diversos metabolitos. El conjunto de estas disciplinas que estudian los diferentes aspectos secuenciales de la funcionalidad celular, conforma la llamada cascada ómica (Santiago-Rodríguez & Hollister, 2021).

1.2.1. Transcriptómica

La transcriptómica es la ciencia ómica que estudia el conjunto de RNAs o transcritos que hay en una muestra en un momento concreto (Koks *et al.*, 2021). A diferencia del genoma, cuya naturaleza es mucho más estática, el transcriptoma cambia constantemente con los diferentes estados de las células. Por ello se estudia cualitativa y cuantitativamente, es decir, se determina qué transcritos hay en la muestra y en qué cantidad (Hasin *et al.*, 2017). La transcriptómica se puede centrar en los distintos tipos de RNA, como son el mRNA, los microRNAs, los long non-coding RNAs, los RNA ribosomales... según se quiera estudiar la expresión, funciones reguladoras o funciones estructurales del transcriptoma (Santiago-Rodriguez & Hollister, 2021; Hasin *et al.*, 2017).

Las tecnologías más habituales para hacer estudios transcriptómicos son microarrays y secuenciación de RNA (RNA-seq). En el caso de los microarrays, requieren de un conocimiento previo para el diseño de las sondas que hibridarán con los transcritos, por lo que sólo se medirán aquellos transcritos que ya se conozcan. No obstante, el análisis es relativamente sencillo e incluso permite el estudio de otros tipos de RNA además del mensajero (Koks *et al.*, 2021). En el caso de RNA-seq, la naturaleza de esta herramienta permite analizar la secuencia de los transcritos, así como el estudio de un mayor rango de RNAs, ya que no está sujeto al diseño de sondas (Santiago-Rodriguez & Hollister, 2021; Koks *et al.*, 2021).

1.2.2. Análisis bioinformáticos de datos de expresión

Los estudios ómicos manejan un volumen muy elevado de datos gracias a las tecnologías que los generan. El análisis e interpretación de estos datos masivos debe ser transparente y reproducible, con el fin de que las conclusiones extraídas tengan validez. Para ello se desarrollan herramientas computacionales que, adaptándose a la naturaleza de los datos y al objetivo del estudio, facilitan los análisis transcriptómicos (Nekrutenko & Taylor, 2012; Lowe *et al.*, 2017).

Para estudiar los datos de expresión es necesario llevar a cabo un preprocesado, en el que se extraen los aspectos más significativos mediante métodos como la corrección del fondo o la normalización. Una vez los datos están preparados, los estudios de expresión se suelen centrar en tres objetivos: descubrimiento de clases, comparación de clases y predicción de clases (Tarca *et al.*, 2006; Lowe *et al.*, 2017). En primer lugar, el descubrimiento de clases consiste en identificar grupos que presentan ciertos aspectos en común dentro de un conjunto de datos. La comparación de clases, por su parte, pretende constatar las diferencias entre dos conjuntos de datos que pueden ser sanos y enfermos u hombres y mujeres, como es el caso del presente trabajo. Por último, la predicción de clases se dedica a desarrollar herramientas de clasificación de muestras en clases conocidas (Tarca *et al.*, 2006).

1.2.3. Metaanálisis

El metaanálisis es una metodología estadística que permite integrar los resultados de distintos estudios que tratan sobre un mismo tema. Su aplicación conjunta incrementa la potencia estadística frente al análisis de los estudios individuales, por lo que aumenta la precisión de las predicciones (Lee, 2015; Lee, 2018). De hecho, el metaanálisis se encuentra en lo alto de la pirámide de la jerarquía de evidencia (figura 1.3), la cual clasifica

las evidencias clínicas en función de los sesgos, de manera que es la evidencia más rotunda (Haidich, 2010).

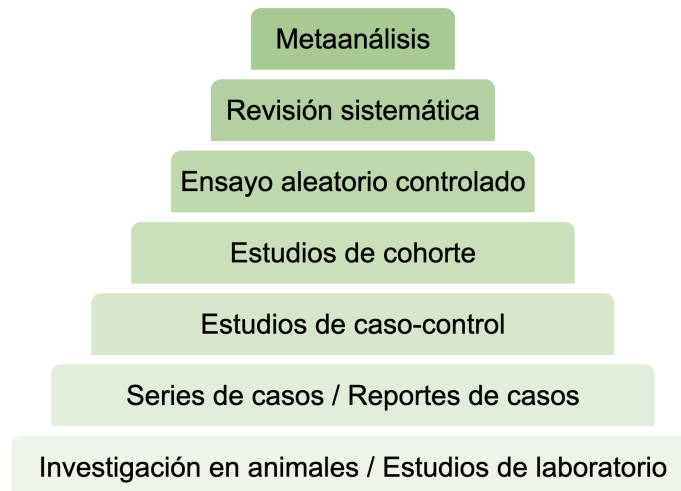


Figura 1.3. Jerarquía de la evidencia (adaptado de Haidich, 2010).

Lee (2018) resume el procedimiento general de los metaanálisis en seis pasos:

1- Seleccionar los estudios relevantes. Es un punto clave, ya que será el punto de partida del análisis, por lo que se deben realizar búsquedas en varias bases de datos y con criterios razonables y establecidos.

2- Examinar la heterogeneidad entre los estudios. Se analizan las posibles diferencias en los resultados de los diferentes estudios.

3- Revisar las diferencias en el metaanálisis al emplear modelos fijos y aleatorios. En los modelos fijos sólo se considera la variación dentro de cada estudio, mientras que en los modelos aleatorios se considera la variación dentro del estudio y también entre ellos. La elección del tipo de modelo dependerá del grado de heterogeneidad.

4- Evaluar las causas de la heterogeneidad. Se lleva a cabo con el fin de asegurar que es viable comparar esos datos, y se puede emplear el análisis de subgrupos, el test de sensibilidad o la metarregresión.

5- Considerar el sesgo de publicación. Éste se presenta como consecuencia del hecho de que los estudios con resultados positivos tienen una mayor posibilidad de publicarse que aquellos que presentan resultados nulos. Se evalúa la precisión o el tamaño de muestra de cada estudio que se incluye para minimizar los efectos del sesgo de publicación en el metaanálisis.

6- Presentar el metaanálisis de acuerdo con las recomendaciones PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Page *et al.*, 2021). Seguir estas indicaciones ayuda a mejorar la calidad de las revisiones sistemáticas y los metaanálisis.

2. Objetivos

El objetivo principal de este trabajo es la identificación y comprensión de los mecanismos moleculares diferenciales por sexo, en la esclerosis múltiple. Este objetivo se alcanzará mediante el cumplimiento de los siguientes objetivos específicos:

1. Revisión sistemática y selección de estudios de esclerosis múltiple con datos de microarrays o secuenciación masiva de microRNAs en los repositorios de datos públicos.
2. Análisis individual de cada estudio seleccionado, incluyendo su procesamiento y un análisis de expresión diferencial.
3. Metaanálisis de microRNAs y de los resultados funcionales que permitirá la detección de biomarcadores específicos por sexo.

3. Materiales y Métodos

El presente trabajo consta en primer lugar, de una revisión sistemática, cuya finalidad es identificar los estudios que presenten conjuntos de datos con la información necesaria para el análisis. Esta revisión sistemática va seguida de un metaanálisis, cuyo objetivo final es identificar nuevos biomarcadores microRNAs en MS con un perfil diferencial por sexo, en el conjunto de estudios seleccionados.

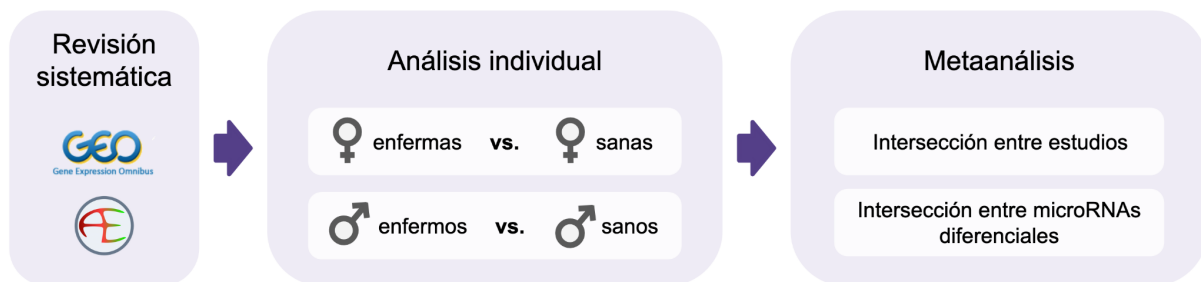


Figura 3.1. Diagrama de las tres fases del trabajo, revisión sistemática, análisis individual y metaanálisis con sus respectivas etapas (♀: mujeres; ♂: hombres).

3.1. Revisión sistemática

De acuerdo con los estándares PRISMA (Page *et al.*, 2021) se llevó a cabo la revisión sistemática. Esta consiste en realizar búsquedas exhaustivas en repositorios con el fin de recopilar conjuntos de datos que cumplan los requisitos para ser incluidos en el posterior metaanálisis. La identificación de estos estudios de entre todos los que aparecen en las búsquedas, se produce aplicando los criterios que se establecen para que actúen como filtros, y así ir descartando los estudios que no son adecuados. Los distintos pasos del proceso general de la revisión sistemática se recogen en el diagrama de flujo PRISMA de la figura 3.2.

El primer paso es realizar las búsquedas de estudios en repositorios (bases de datos). Las directrices PRISMA (Page *et al.*, 2021) hablan del empleo de más de una fuente para buscarlos, ya que de esta manera se abarcan un mayor número de entradas y puede que algunas de ellas sólo se encuentren en uno de los repositorios. Por ello, las búsquedas se realizaron en las bases de datos Gene Expression Omnibus (GEO) (Barrett *et al.*, 2019) y ArrayExpress (Athar *et al.*, 2019). Las palabras clave utilizadas fueron “multiple sclerosis” y “ms”. Los filtros de las búsquedas, que permiten acotar los resultados, fueron aplicados en medida de las posibilidades que ofrece cada base de datos. Con todo ello, las búsquedas realizadas fueron:

- *ms*[All Fields] AND (“gse”[Filter] AND “Homo sapiens”[Organism] AND (“Non-coding RNA profiling by array”[Filter] OR “Non-coding RNA profiling by high throughput sequencing”[Filter])) en la base de datos GEO y en las fechas 8/2/2022 y 9/2/2022.
- (“multiple sclerosis”[MeSH Terms] OR multiple sclerosis[All Fields]) AND (“gse”[Filter] AND “Homo sapiens”[Organism] AND (“Non-coding RNA profiling by array”[Filter] OR “Non-coding RNA profiling by high throughput sequencing”[Filter])) en la base de datos GEO y en la fecha 11/2/2022.

- *Search for "multiple sclerosis" Filtered by organism Homo sapiens, experiment type "rna assay", experiment type "array assay"* en la base de datos ArrayExpress y en la fecha 25/2/2022.
- *Search for "multiple sclerosis" Filtered by organism Homo sapiens, experiment type "rna assay", experiment type "sequencing assay"* en la base de datos ArrayExpress y en la fecha 6/3/2022.
- *Search for "ms" Filtered by organism Homo sapiens, experiment type "rna assay", experiment type "array assay"* en la base de datos ArrayExpress y en la fecha 6/3/2022.
- *Search for "ms" Filtered by organism Homo sapiens, experiment type "rna assay", experiment type "sequencing assay"* en la base de datos ArrayExpress y en la fecha 6/3/2022.

Una vez recopilados todos los estudios resultantes de las búsquedas, el primer paso fue eliminar los duplicados. Es común que estos aparezcan, ya que se emplearon varias palabras clave que hacen referencia a la misma enfermedad, y además existen conjuntos de datos que están presentes en ambas bases de datos.

A continuación, se aplicaron los criterios básicos para excluir los estudios fácilmente descartables. Estos criterios básicos eran tres: (I) no ser un estudio centrado en MS, (II) no analizar datos de microRNAs y (III) estudiar muestras que no procedían directamente de organismos humanos, como por ejemplo líneas celulares o ratones.

Seguidamente se establecieron criterios más específicos para seguir descartando estudios que no cumplían los requisitos para ser incluidos en el metaanálisis. En este caso se excluyeron los conjuntos de datos que (IV) no contenían información del sexo del paciente, (V) estudiaban pacientes de un solo sexo y/o (VI) evaluaban el efecto de un tratamiento sobre el transcriptoma de los pacientes.

Por último, los estudios resultantes del cribado se sometieron a un análisis exploratorio más exhaustivo para determinar si el tipo de procesamiento de los datos era adecuado para su integración posterior mediante el metaanálisis.

Esta exploración incluyó la revisión del archivo con los datos normalizados por estudio. Simultáneamente se comprobó que las muestras de pacientes de cada sexo, descritas en el repositorio público, coincidían con la descripción muestral de la publicación en el caso de que la hubiera, con el fin de verificar que los datos habían sido interpretados correctamente. Para finalizar este análisis, se comprobó la disponibilidad de la herramienta GEO2R implementada en la base de datos GEO para cada uno de los conjuntos de datos. De esta forma, se descartaron todos aquellos estudios que no eran adecuados para el análisis integrativo y sólo fueron seleccionados aquellos que cumplieron todos los criterios descritos.

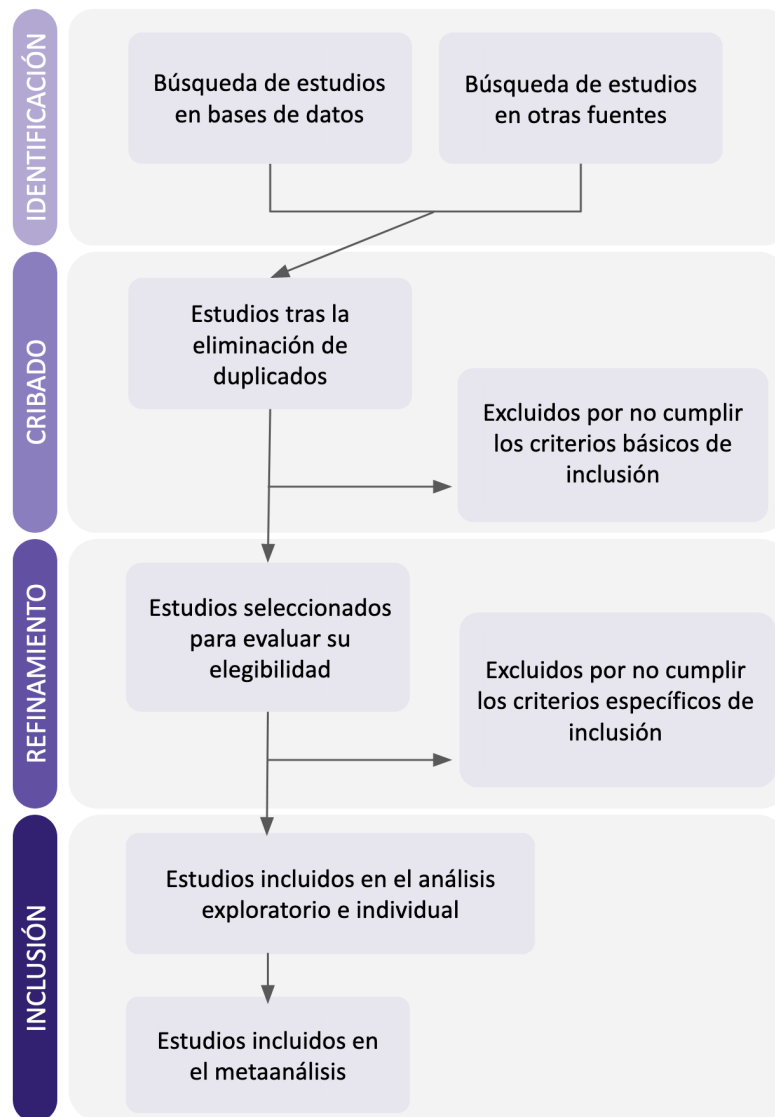


Figura 3.2. Diagrama de flujo de revisión sistemática según los estándares PRISMA (adaptado de Page et al., 2021).

3.2. Análisis individual de los estudios

En cada uno de los estudios disponemos de muestras correspondientes a 4 grupos experimentales: mujeres enfermas, mujeres sanas, hombres enfermos y hombres sanos. Se realizó un análisis descriptivo para conocer la distribución de las muestras en los grupos experimentales indicados. Seguidamente, se empleó la herramienta web GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) que permite comparar los grupos de muestras seleccionados dentro de cada estudio. Esta herramienta emplea los paquetes GEOquery y limma del lenguaje de programación R (R Core Team, 2021) para identificar genes expresados diferencialmente.

Las comparaciones que se llevaron a cabo fueron: mujeres enfermas vs. mujeres sanas y hombres enfermos vs. hombres sanos. Para cada una de estas dos comparaciones se obtuvo una tabla que indicaba el identificador del microRNA expresado diferencialmente junto a su P-valor asociado, así como los valores estadísticos t, B y logFC. Adicionalmente se generaron gráficos que permitieron la visualización de los resultados, como el volcano

plot para resumir el nivel de expresión diferencial, el box plot para describir la distribución de los valores de expresión, o el gráfico de diferencia media.

Se seleccionaron los microRNAs con expresión diferencial significativa, es decir con un P-valor < 0,05. A continuación se separaron aquellos que estaban sobreexpresados en esclerosis múltiple (valor de logFC negativo) de los que estaban inhibidos en los pacientes (valor de logFC positivo). De esta forma, para cada estudio se obtuvieron cuatro listas de identificadores de microRNAs significativos, cada una de ellas se corresponde con un perfil de expresión diferencial que se considera en el presente estudio:

- Sobreexpresados en mujeres con MS
- Inhibidos en mujeres con MS
- Sobreexpresados en hombres con MS
- Inhibidos en hombres con MS

3.3. Metaanálisis

Una vez obtenidos los resultados del análisis individual de cada estudio, se desarrollaron estrategias de análisis integrativo para combinar los resultados en su conjunto. Se llevaron a cabo una serie de comparaciones con el objetivo de buscar coincidencias entre los distintos estudios que ratificaran los resultados.

En un primer análisis se compararon mediante diagramas de Venn cada una de las listas correspondientes a una clase de expresión diferencial procedentes de cada estudio como se ilustra en la figura 3.3, de manera que se compararon las listas de microRNAs sobreexpresados en mujeres enfermas de los cuatro estudios para buscar coincidencias entre ellos (análisis de intersección). Esto mismo se repitió para los microRNAs inhibidos, e igualmente en el caso de los hombres. Estas cuatro comparaciones tienen el objetivo de constatar si hay microRNAs significativos que aparecen en varios de los estudios seleccionados. La herramienta empleada para obtener estos diagramas de Venn fue Venny 2.1.0 (Oliveros, 2007-2015), un recurso web diseñado para comparar listas.

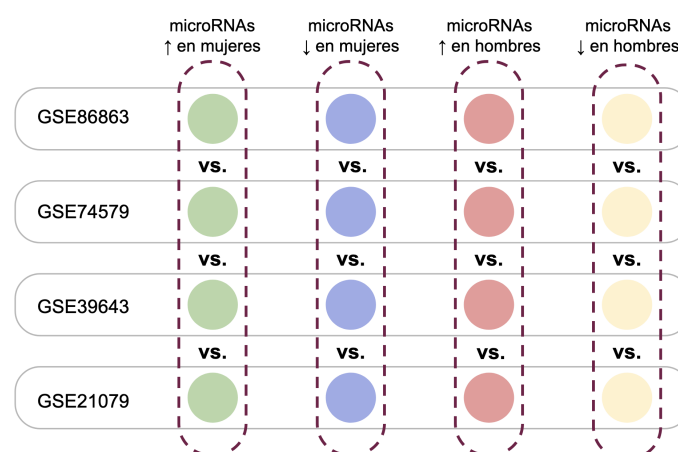


Figura 3.3. Descripción de las cuatro comparaciones entre los estudios correspondientes al primer análisis (↑ para microRNAs sobreexpresados y ↓ para inhibidos, los datos comparados se rodean con líneas discontinuas).

En un segundo análisis se combinaron los datos de los diferentes estudios de forma aditiva (cada estudio aporta sus microRNAs significativos sean coincidentes o no con otros estudios), por lo que se obtuvieron las cuatro listas que englobaban todos los microRNAs sobreexpresados en mujeres enfermas, los inhibidos en mujeres enfermas, los sobreexpresados en hombres enfermos y los inhibidos en hombres enfermos, como se observa en la figura 3.4. Con las listas que reunían la información de los cuatro estudios seleccionados, se generó otro diagrama de Venn, con el objetivo de identificar los microRNAs que presentaban determinados perfiles de expresión en función del sexo, así como los comunes a hombres y mujeres

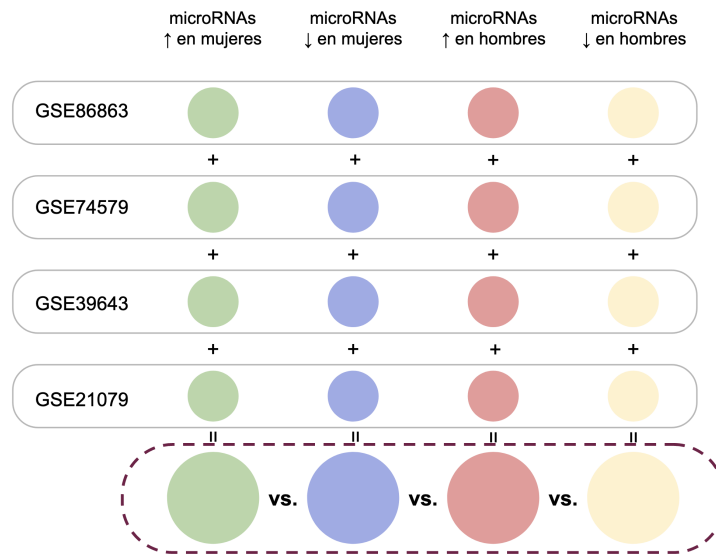


Figura 3.4. Descripción de la comparación realizada en el segundo análisis, entre cada perfil de expresión sumando los datos aportados por cada estudio (↑ para microRNAs sobreexpresados y ↓ para inhibidos, los datos comparados se rodean con líneas discontinuas).

Para conocer el papel biológico de los microRNAs identificados con las estrategias de integración, se empleó la herramienta Diana tools (Vlachos *et al.*, 2015a), que permitió detectar qué funciones biológicas estaban asociadas a los biomarcadores seleccionados. Este enriquecimiento funcional se lleva a cabo sobre los genes que constituyen las dianas de los microRNAs. El análisis requería emplear las formas maduras de los microRNAs, por lo que se recurrió a miRBase (Griffiths-Jones *et al.*, 2006) con el fin de recopilar todos los posibles microRNAs maduros. Las funciones biológicas significativas se representaron mediante heatmaps, analizando los procesos biológicos con términos GO (Gene Ontology) y con rutas metabólicas KEGG (Kyoto Encyclopedia of Genes and Genomes).

4. Resultados

4.1. Revisión sistemática

El resultado del proceso de revisión sistemática se presenta en la figura 4.1. Tras la búsqueda realizada se identificaron un total de 444 estudios entre las bases de datos GEO y ArrayExpress. Después de la eliminación de los estudios duplicados, el número de trabajos se redujo a 361. Este número resulta elevado principalmente a raíz de la búsqueda de la palabra clave “ms” en ArrayExpress. Las opciones de filtrado en esta base de datos son limitadas, por lo que aparecen muchos estudios que no están realmente relacionados con la esclerosis múltiple.

Para continuar con el cribado, se excluyeron 340 estudios que no cumplían los criterios básicos de inclusión. La mayoría de estudios eliminados en este paso (257) no estaban relacionados con MS, aunque también se descartaron estudios que no incluían información de microRNAs o que no empleaban muestras procedentes directamente de organismos humanos.

De los 21 estudios restantes, otros 14 fueron descartados porque, bien no presentaban información sobre el sexo de los pacientes, bien no incluían pacientes de ambos sexos, o bien estudiaban los efectos de distintos tratamientos sobre los pacientes.

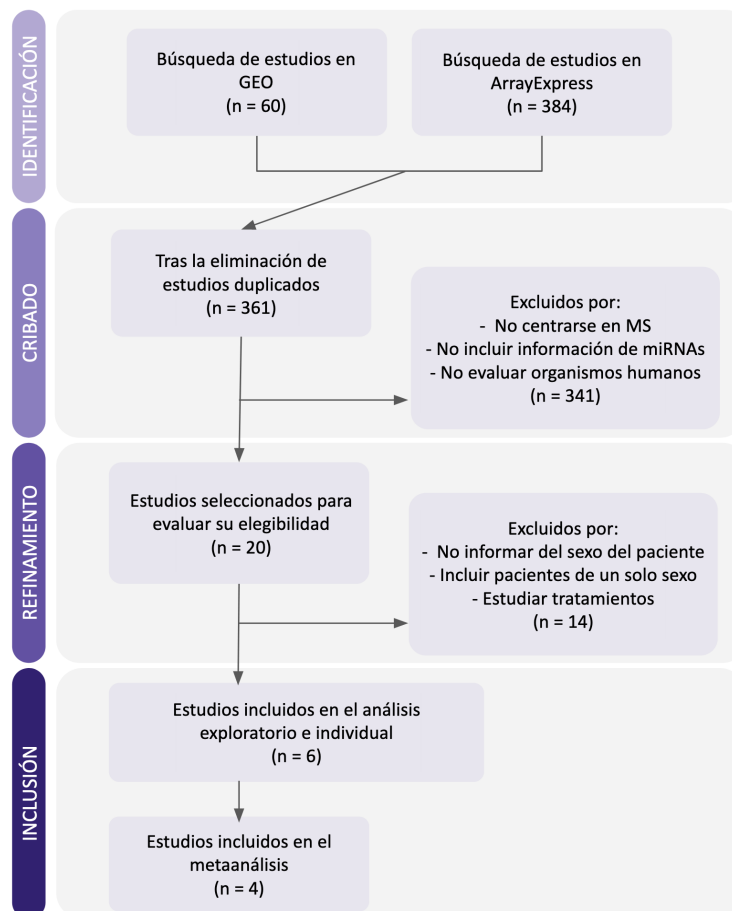


Figura 4.1. Resultados de la revisión sistemática de estudios de expresión en esclerosis múltiple, mediante la aplicación de los estándares de PRISMA.

Finalmente, resultaron 6 estudios que se plasman en la tabla 4.1, los cuales fueron evaluados mediante un análisis individual para su inclusión en el metaanálisis. De estos 6, se excluyeron 2. Uno de ellos (E-MTAB-359) fue descartado porque no incluía el archivo de datos normalizados (imprescindible para poder comparar ese conjunto de datos con el resto). El segundo estudio (GSE124900) se descartó por presentar el archivo de datos normalizados en un formato incompatible con el uso de la herramienta de análisis GEO2R, implementada en la interfaz de GEO y que se usaría para el análisis individual de cada estudio.

Tabla 4.1. Selección final de estudios para el metaanálisis. Se incluyen identificador, base de datos de procedencia, plataforma de estudio transcriptómico y referencia.

Identificador	Base de datos	Plataforma	Referencia
GSE86863	GEO	GPL18941 3D-Gene Human miRNA V20_1.0.0	Sin referencia
GSE74579	GEO	GPL8786 [miRNA-1] Affymetrix Multispecies miRNA-1 Array	Muñoz-Culla et al. (2016)
GSE39643	GEO	GPL15847 NanoString nCounter Human miRNA assay	Butovsky et.al (2012)
GSE21079	GEO	GPL8178 Illumina Human v1 MicroRNA expression beadchip	Cox et al. (2010)

4.2. Análisis individual de los estudios

En primer lugar, se identificaron los cuatro grupos experimentales en cada estudio para realizar un recuento de individuos por grupo (tabla 4.2). La distribución de muestras por grupo no está balanceada, aunque sí hay representación de sujetos en todos ellos, asegurando que se pueda llevar a cabo la integración de los resultados individuales.

Tabla 4.2. Recuento de individuos pertenecientes a los grupos experimentales en los estudios seleccionados para el metaanálisis.

Estudio	Mujeres con MS	Mujeres sanas	Hombres con MS	Hombres sanos	Individuos totales
GSE86863	2	2	2	2	8
GSE74579	28	14	20	10	72
GSE39643	4	4	4	4	16
GSE21079	40	21	19	16	96
Total	74	41	45	32	192

Abreviaturas - MS: esclerosis múltiple.

A continuación y en cada uno de los estudios, se analizaron los microRNAs expresados diferencialmente mediante GEO2R. Se compararon por un lado las mujeres con MS con las mujeres sanas, y por otro lado los hombres con MS y los hombres sanos. De cada

comparación se obtuvieron dos listas de microRNAs que presentan expresión diferencial significativa, una con aquellos sobreexpresados en los pacientes y otra con aquellos inhibidos. Estos constituyen los cuatro perfiles de microRNAs diferenciales que se estudian. Los resultados de estas comparaciones se recogen en la tabla 4.3.

Tabla 4.3. Número de microRNAs significativos con sobreexpresión (\uparrow) o expresión inhibida (\downarrow) en cada estudio.

Estudio	microRNAs \uparrow mujeres	microRNAs \downarrow mujeres	microRNAs \uparrow hombres	microRNAs \downarrow hombres
GSE86863	1	12	5	11
GSE74579	22	8	10	14
GSE39643	39	12	24	10
GSE21079	17	45	15	10

4.3. Metaanálisis

El primer análisis llevado a cabo sobre las listas de microRNAs consistía en comparar los resultados de los distintos estudios para cada clase de expresión diferencial. Los resultados se presentan en diagramas de Venn en la figura 4.2. Se observa que las intersecciones entre los distintos estudios son pequeñas y no incluyen todos los estudios.

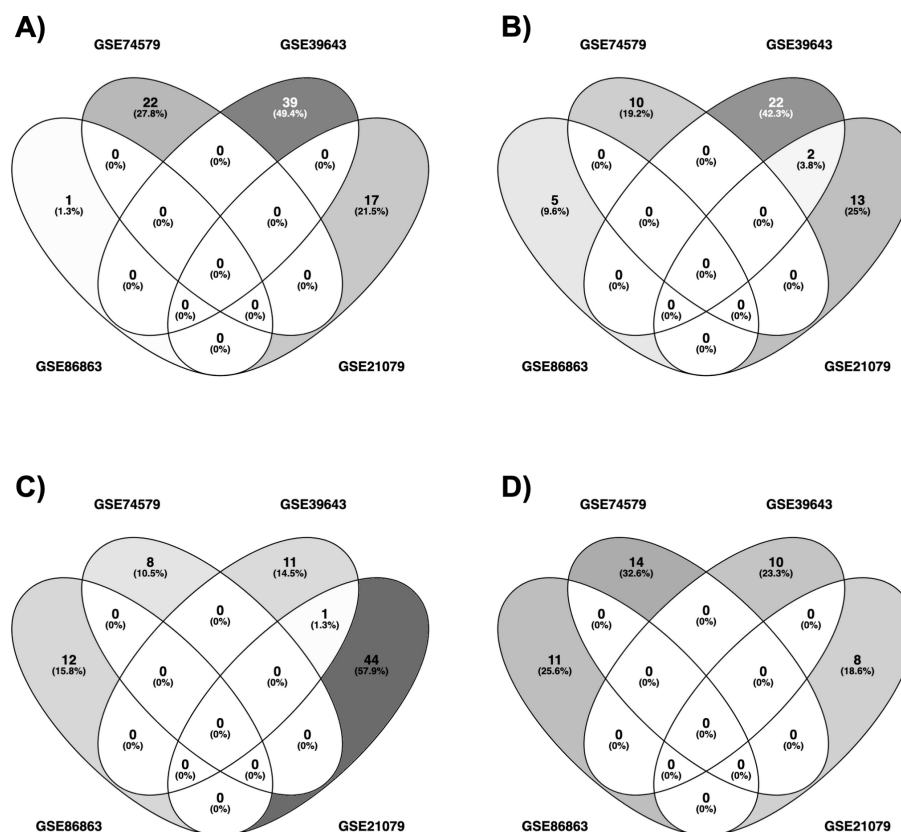


Figura 4.2. Diagramas de Venn presentando las intersecciones entre los estudios GSE86863, GSE74579, GSE39643 y GSE21079 en microRNAs sobreexpresados en

mujeres con MS (A), *microRNAs* sobreexpresados en hombres con MS (B), *microRNAs* inhibidos en mujeres con MS (C) y *microRNAs* inhibidos en hombres con MS.

A continuación, se llevó a cabo un segundo análisis, esta vez sumando los *microRNAs* significativos de cada estudio, para cada perfil de *microRNAs* con expresión diferencial. De esta forma, se compararon los cuatro perfiles de *microRNAs* diferenciales. Es decir, los sobreexpresados e inhibidos tanto en mujeres como en hombres enfermos aportados por cada uno de los estudios. El diagrama de Venn que ilustra esta comparación se presenta en la figura 4.3. Las intersecciones clave que aporta este diagrama son (1) las que identifican *microRNAs* con perfiles opuestos en cada sexo (entre el amarillo y el verde, y entre el azul y el rojo), y (2) aquellas que indican los *microRNAs* con el mismo perfil en los dos sexos (entre el azul y el amarillo, y entre el verde y el rojo), que serían marcadores de la enfermedad independientes del sexo.

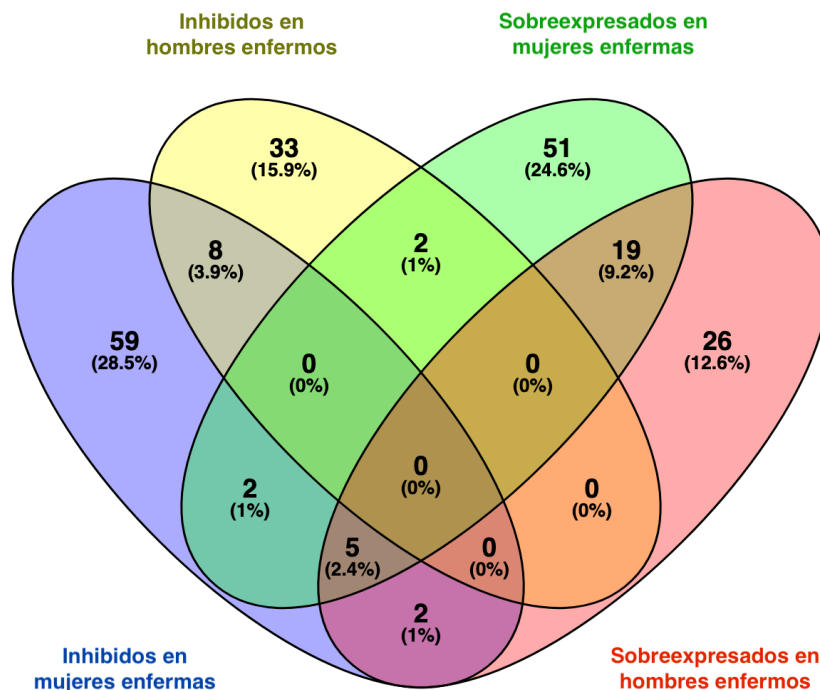


Figura 4.3. Diagrama de Venn que compara los *microRNAs* con los distintos perfiles de expresión diferencial, señalando las intersecciones entre ellos. Se identifican las intersecciones que señalan los *microRNAs* con perfiles opuestos y con perfiles iguales en hombres y mujeres.

Entre los *microRNAs* con perfil opuesto en ambos sexos se encuentran 2 con una expresión baja en hombres y elevada en mujeres, y otros 2 con alta expresión en hombres y baja en mujeres. Por otro lado, aparecen 8 con expresión baja y 19 con expresión elevada en todos los pacientes, independientemente de su sexo. Todos estos *microRNAs* se registraron mediante sus identificadores (tabla 4.4) con el fin de llevar a cabo su análisis funcional.

Tabla 4.4. Resumen de los microRNAs resultantes de la intersección entre las clases de expresión diferencial (↑: sobreexpresión; ↓: inhibición).

Perfil de expresión	Identificadores
↑ Mujeres / ↓ Hombres	hsa-mir-181a-1 // hsa-mir-181a-2 hsa-mir-29c
↓ Mujeres / ↑ Hombres	hsa-mir-1228 hsa-miR-196a
↑ Mujeres / ↑ Hombres	hsa-let-7b hsa-miR-26a hsa-miR-361-5p hsa-miR-103 hsa-miR-423-5p hsa-miR-197 hsa-miR-664 hsa-miR-24 hsa-miR-320c hsa-miR-423-3p hsa-miR-1274b hsa-miR-720 hsa-miR-1260 hsa-miR-1274a hsa-miR-146a hsa-miR-532-3p hsa-miR-425 hsa-miR-181a hsa-miR-768-3p:11.0
↓ Mujeres / ↓ Hombres	hsa-miR-142-3p hsa-miR-450a hsa-miR-598 hsa-miR-649 hsa-miR-383 hsa-miR-362-3p hsa-miR-579 hsa-miR-633

El análisis funcional se realizó mediante la herramienta Diana tools, que permite tanto el enriquecimiento funcional con rutas KEGG como con términos GO. Para cada perfil de expresión se unieron los microRNAs asociados, y a continuación se evaluaron las rutas metabólicas (KEGG) y los procesos biológicos (GO) relacionados con ellos. Cabe señalar la necesidad de emplear las formas maduras de los microRNAs para el análisis, así como la exclusión de aquellos microRNAs que no constan en la base de datos de Diana tools DIANA-TarBase v7.0 (Vlachos *et al.*, 2015b) que fueron 4, todos ellos pertenecientes a el perfil de elevada expresión en ambos sexos.

Para visualizar las funciones biológicas enriquecidas se emplearon los gráficos heatmap, de manera que para cada perfil de expresión (opuestos e iguales) se elaboraron dos heatmaps correspondientes a las rutas KEGG y los términos GO. De esta manera, se identificaron los

enriquecimientos que indicaban las funciones que se dan en exceso en caso de la sobreexpresión, y las que se dan por debajo de los valores normales en caso de inhibición.

En la figura 4.4 se muestran los heatmaps más informativos relacionados con el análisis de enriquecimiento funcional de rutas KEGG y términos GO. Todos los gráficos se encuentran con mayor detalle en el anexo. Los enriquecimientos de rutas metabólicas y procesos biológicos con un mayor poder estadístico de cada perfil de expresión se describen en la tabla 4.5.

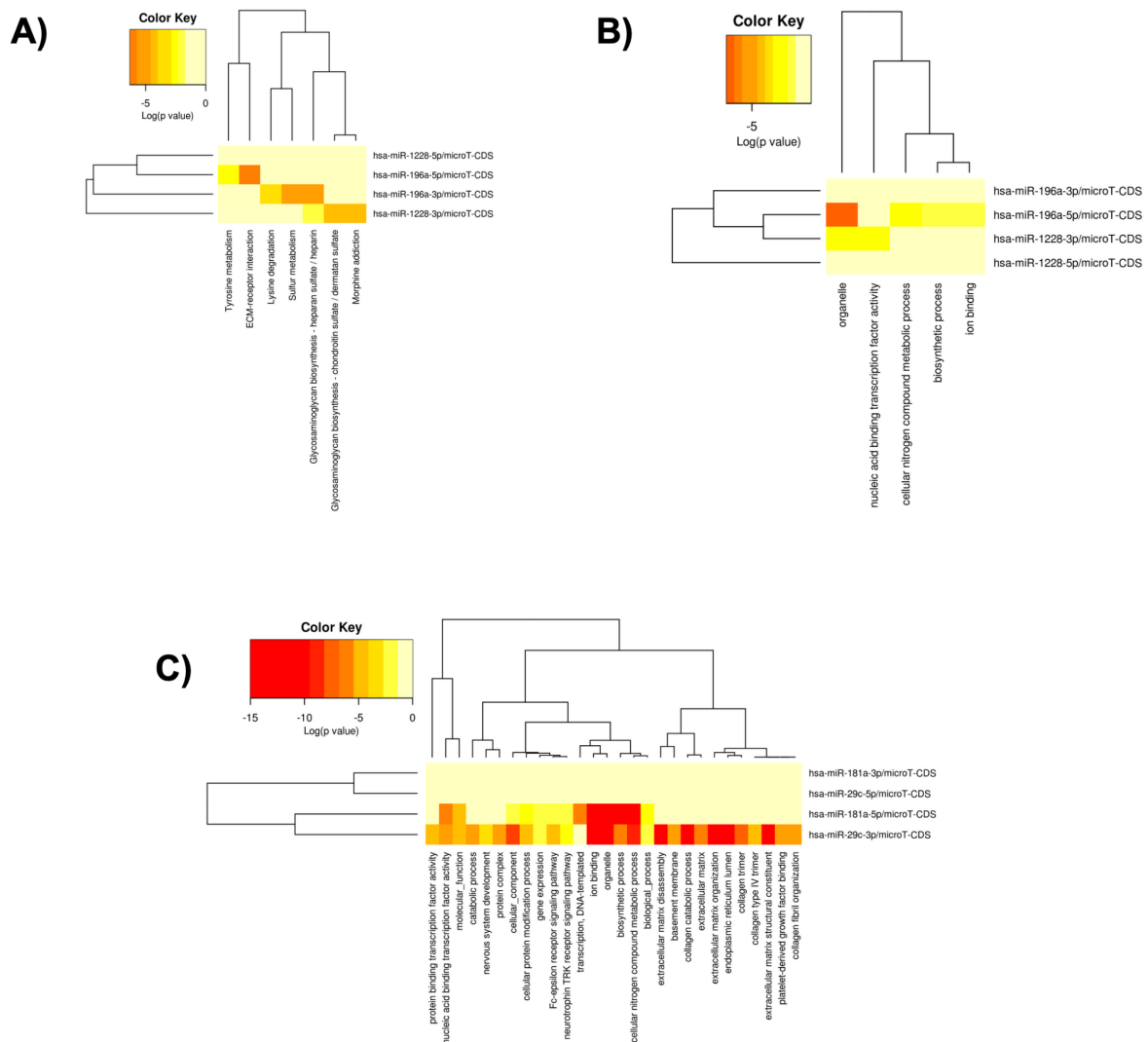


Figura 4.4 Selección de heatmaps en representación de los resultados del enriquecimiento funcional. Los microRNAs con perfil de expresión bajo en mujeres y alto en hombres presentan enriquecimientos con menor potencia estadística, tanto en rutas KEGG (A) como en términos GO (B); mientras que los demás, como los de perfil de expresión alto en mujeres y bajo en hombres (C) presentan enriquecimientos más claros.

Tabla 4.5. Resumen de las rutas KEGG y términos GO con el enriquecimiento más significativo de cada perfil de expresión (↑: sobreexpresión; ↓: inhibición).

Perfil	Análisis	Ruta / PB	P-valor	Nº genes	Nº microRNAs
↑ Mujeres ↓ Hombres	KEGG	Interacción ECM-receptor	0	18	1
	GO	Procesado metabólico de compuestos nitrogenados	0	288	2
		Unión de iones	0	367	2
		Orgánulo	0	532	2
↓ Mujeres ↑ Hombres	KEGG	Biosíntesis de glicosaminoglicanos	7.341137E-05	3	2
	GO	Orgánulo	2.37102E-08	283	2
↑ Mujeres ↑ Hombres	KEGG	Interacción ECM-receptor	5.324272E-10	10	2
	GO	Actividad de unión de TF a ácidos nucleicos	0	323	8
		Ruta de señalización del receptor Fc-ε	0	83	12
		Nucleoplasma	0	349	8
↓ Mujeres ↓ Hombres	KEGG	Metabolismo de xenobióticos por el citocromo P450	0	3	2
	GO	Ruta de señalización del receptor TRK neurotrofina	0	71	5
		Procesos de modificación de proteínas	0	478	6
		Orgánulo	0	1884	8

Abreviaturas - PB: proceso biológico; : matriz extracelular; TF: factor de transcripción.

5. Discusión

La MS es una enfermedad que presenta claras diferencias entre pacientes en función del sexo. Esto se puede observar en la incidencia incrementada en mujeres y en la progresión acelerada y más severa en hombres (Voskuhl, 2020). Estas diferencias clínicas deben llevar asociadas sus diferencias a nivel molecular, y conocerlas en profundidad puede ayudar en el diagnóstico y en la toma de decisiones a la hora de aplicar tratamientos. Este último punto es clave, ya que permite predecir la eficacia de los tratamientos en lugar de aplicarlos a ciegas, considerando los efectos secundarios que presentan.

Durante décadas se ha tomado el modelo masculino como norma general en la caracterización de enfermedades, y es cierto que en la actualidad a nivel clínico, este sesgo se ha empezado a subsanar. Sin embargo, aún es habitual que se ignore la variable del sexo en los diseños experimentales de los ensayos biomédicos, dejando de lado la información clave que esta variable puede aportar (Clayton & Collins, 2014). En el caso de la MS, junto con otras enfermedades autoinmunes y neurodegenerativas, el estudio de la influencia de la componente sexo a nivel molecular es de gran importancia para alcanzar una mayor comprensión sobre las diferencias entre sexos que se manifiestan a nivel clínico (Moulton, 2018; Vegeto *et al.*, 2020).

Dichas deficiencias experimentales son palpables en este estudio, ya que en la fase de revisión sistemática regida por la normativa PRISMA, tan solo 63 de los 187 estudios centrados en MS incluían la variable del sexo en sus datos. Por tanto, no fue posible la reutilización de un volumen elevado de datos, que podrían haber sumado poder estadístico a los resultados. Además, de los 6 estudios que contenían la información necesaria, 2 tuvieron que ser excluidos (el 33%) por motivos relacionados con la forma de publicar los datos. Por ello es importante hacer hincapié en la relevancia de la estandarización y el consenso en el formato a la hora de publicar los datos.

Al empezar a analizar los estudios seleccionados mediante GEO2R en el análisis individual, se observa que el número de muestras que componen cada grupo experimental no está equilibrado, y a pesar de que esta situación no es la ideal, el tamaño muestral permite llevar a cabo el análisis. Las comparaciones entre pacientes y sanos de cada sexo, proporcionó para cada uno de los estudios dos conjuntos de microRNAs con expresión diferencial significativa, por un lado en mujeres y por otro en hombres.

En estos conjuntos de microRNAs con expresión diferencial procedentes del análisis individual se basó el metaanálisis, que estaba compuesto de dos análisis de intersección. La primera intersección incluía 4 comparaciones, una para cada perfil de expresión diferencial (sobrexpresados e inhibidos tanto en hombres como en mujeres). Se hizo una comparación entre estudios con el objetivo de encontrar coincidencias entre los microRNAs aportados por cada estudio. No obstante, no se halló ninguna intersección entre los cuatro estudios en ninguna de las comparaciones.

Alternativamente, para el segundo análisis de intersección se sumaron los microRNAs de cada perfil de expresión aportados por cada estudio, por lo que se hizo una comparación integrativa que permitió identificar microRNAs que presentaban perfiles iguales u opuestos en cada sexo. Se obtuvieron 2 microRNAs con expresión elevada en mujeres y baja en hombres y otros 2 con expresión reducida en mujeres y alta en hombres. A su vez se

identificaron 19 microRNAs con expresión elevada y 8 con expresión reducida en ambos sexos. Estos dos últimos grupos de microRNAs constituyen un conjunto de marcadores de la enfermedad independientes del sexo del paciente; mientras que los de perfiles opuestos señalan las posibles diferencias a nivel molecular entre pacientes hombres y mujeres. Con todos ellos se realizó un análisis funcional de rutas KEGG y de términos GO con el fin de identificar rutas metabólicas y procesos biológicos enriquecidos en cada perfil de expresión. Ello se llevó a cabo con la herramienta Diana tools empleando las formas maduras de los microRNAs, y se representaron los resultados mediante heatmaps.

En cuanto a los perfiles de expresión opuestos en los dos sexos, encontramos por un lado 2 microRNAs con funciones aumentadas en mujeres que a la vez están disminuidas en hombres. Entre estas funciones destacan la ruta metabólica relacionada con la interacción de la matriz extracelular (ECM) y sus receptores, así como los procesos biológicos de compuestos nitrogenados, unión de iones y orgánulos. Estas asociaciones pueden indicar una desregulación de las moléculas receptoras ya que hay mayor interacción con la ECM y también con los iones, lo que puede derivar en un incremento de las señalizaciones a nivel celular en las mujeres y un descenso en hombres.

Por otro lado, los 2 microRNAs con funciones disminuidas en mujeres y aumentadas en hombres se asocian a la ruta de biosíntesis de glicosaminoglicanos y al proceso biológico de orgánulo. Es cierto que este perfil es el que presenta resultados con menor valor estadístico, puede que sea el motivo de el número reducido de funciones relacionadas. Asimismo la función enriquecida de orgánulo aparece en prácticamente todos los heatmaps, esto puede deberse a que es un término que abarca una amplia cantidad de moléculas biológicas, por lo que aporta poca información.

En relación a los perfiles de expresión concordantes en ambos sexos, se encontraron por un lado 19 microRNAs con sobreexpresión respecto a los individuos sanos. Estos estaban asociados a la ruta metabólica de interacción de la ECM con sus receptores; así como a los procesos biológicos de actividad de unión de factores de transcripción (TF), ruta de señalización del receptor Fc- ϵ y nucleoplasma. Cabe destacar la elevada señalización del receptor de la parte constante de anticuerpos, que sin duda está relacionada con el componente autoinmune de la enfermedad. Así como la mayor actividad de unión de los TF, que lleva a una mayor expresión génica y por tanto puede contribuir a un estado de desregulación.

En el caso de los 8 microRNAs con baja expresión en ambos sexos, la ruta asociada que destaca es el metabolismo de xenobióticos por el citocromo P450; mientras que los procesos biológicos en este caso reducidos son la ruta de señalización del receptor TRK neurotrofina, los procesos de modificación de proteínas y orgánulos. Una actividad baja de modificación de proteínas definitivamente altera las funciones celulares, además una alteración en la señalización de la neurotrofina comporta alteraciones en el desarrollo, mantenimiento y función de las neuronas (Huang & Reichardt, 2001), lo que puede contribuir a la enfermedad.

Los resultados obtenidos en este estudio ratifican que las diferencias de sexo a nivel clínico de la MS vienen asociadas a diferencias a nivel molecular. En concreto, se confirma el papel que juegan los microRNAs en la regulación de la expresión génica de los pacientes, que presentando perfiles de expresión diferenciales respecto a individuos sanos, alteran la

homeóstasis y contribuyen al desencadenamiento y desarrollo de la enfermedad. No obstante, para que estas diferencias puedan ser estudiadas en mayor profundidad, es necesario que la variable del sexo se estudie en los experimentos, ya que permitirá alcanzar una mayor comprensión sobre el origen molecular de las evidentes diferencias clínicas entre sexos que manifiesta la MS. Además, estas diferencias pueden complementar el diagnóstico de la enfermedad, que actualmente se realiza por MRI, ya que se podría establecer un perfil molecular en función de las características del paciente y así hacer una identificación más rápida, sensible y selectiva. Esto va ligado a los beneficios que puede aportar el conocimiento del perfil molecular a la elección del tratamiento más adecuado para el paciente desde un inicio, es decir aplicar la medicina personalizada.

6. Conclusiones

1. La consideración de la variable del sexo de los individuos es un punto clave a mejorar en el diseño experimental de los ensayos biomédicos, dejando aún por conocer información molecular relevante en el estudio de enfermedades.
2. La revisión sistemática y el análisis individual evidencian la falta de consenso y estandarización a la hora de publicar los datos de los estudios, lo que impide su reciclado para obtener más información a partir de ellos.
3. Las técnicas de metaanálisis aplicadas a estudios de expresión de microRNAs permiten integrar la información de diversos grupos de datos ofreciendo una caracterización global de su papel regulatorio.
4. Las herramientas web como GEO2R, Venny y Diana tools hacen más accesibles los análisis y las bases de datos, facilitando la extracción de información de los conjuntos de datos.
5. La identificación de microRNAs expresados diferencialmente en MS, y por tanto las diferencias en la regulación de la expresión génica entre hombres y mujeres, demuestran la necesidad de estudiar en profundidad los mecanismos moleculares enlazados a las diferencias clínicas entre sexos; todo ello con el objetivo de desarrollar la medicina personalizada.

7. Bibliografía

- ATHAR, A.; FÜLLGRABE, A.; GEORGE, N.; IQBAL, H.; HUERTA, L.; ALI, A.; SNOW, C.; FONSECA, N.A.; PETRYSZAK, R.; PAPTAEODOROU, I.; SARKANS, U.; BRAZMA, A. (2019). ArrayExpress update—from bulk to single-cell expression data. *Nucleic Acids Research*, 47(D1), D711–D715. <https://doi.org/10.1093/nar/gky964>
- BARRETT, T., WILHITE, S. E., LEDOUX, P., EVANGELISTA, C., KIM, I. F., TOMASHEVSKY, M., MARSHALL, K. A., PHILLIPPY, K. H., SHERMAN, P. M., HOLKO, M., YEFANOV, A., LEE, H., ZHANG, N., ROBERTSON, C. L., SEROVA, N., DAVIS, S., & SOBOLEVA, A. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*, 41(Database issue), D991–D995. <https://doi.org/10.1093/nar/gks1193>
- BJORNEVIK, K., CORTESE, M., HEALY, B. C., KUHLE, J., MINA, M. J., LENG, Y., ELLEDGE, S. J., NIEBUHR, D. W., SCHER, A. I., MUNGER, K. L., & ASCHERIO, A. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (New York, N.Y.)*, 375(6578), 296–301. <https://doi.org/10.1126/science.abj8222>
- BUTOVSKY, O., SIDDIQUI, S., GABRIELY, G., LANSER, A. J., DAKE, B., MURUGAIYAN, G., DOYKAN, C. E., WU, P. M., GALI, R. R., IYER, L. K., LAWSON, R., BERRY, J., KRICHEVSKY, A. M., CUDKOWICZ, M. E., & WEINER, H. L. (2012). Modulating inflammatory monocytes with a unique microRNA gene signature ameliorates murine ALS. *The Journal of clinical investigation*, 122(9), 3063–3087. <https://doi.org/10.1172/JCI62636>
- CLAYTON, J. A., & COLLINS, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature*, 509(7500), 282–283. <https://doi.org/10.1038/509282a>
- COX, M. B., CAIRNS, M. J., GANDHI, K. S., CARROLL, A. P., MOSCOVIS, S., STEWART, G. J., BROADLEY, S., SCOTT, R. J., BOOTH, D. R., LECHNER-SCOTT, J., & ANZGENE MULTIPLE SCLEROSIS GENETICS CONSORTIUM (2010). MicroRNAs miR-17 and miR-20a inhibit T cell activation genes and are under-expressed in MS whole blood. *PLoS one*, 5(8), e12132. <https://doi.org/10.1371/journal.pone.0012132>
- DOSHI, A., & CHATAWAY, J. (2016). Multiple sclerosis, a treatable disease. *Clinical medicine (London, England)*, 16(Suppl 6), s53–s59. <https://doi.org/10.7861/clinmedicine.16-6-s53>
- GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A., & ENRIGHT, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research*, 34(Database issue), D140–D144. <https://doi.org/10.1093/nar/gkj112>
- Haidich A. B. (2010). Meta-analysis in medical research. *Hippokratia*, 14(Suppl 1), 29–37.
- HASIN, Y., SELDIN, M., & LUSIS, A. (2017). Multi-omics approaches to disease. *Genome biology*, 18(1), 83. <https://doi.org/10.1186/s13059-017-1215-1>
- HOUEN, G., TRIER, N. H., & FREDERIKSEN, J. L. (2020). Epstein-Barr Virus and Multiple Sclerosis. *Frontiers in immunology*, 11, 587078. <https://doi.org/10.3389/fimmu.2020.587078>

- HUANG, E. J., & REICHARDT, L. F. (2001). Neurotrophins: roles in neuronal development and function. *Annual review of neuroscience*, 24, 677–736. <https://doi.org/10.1146/annurev.neuro.24.1.677>
- HRASTELJ, J., & ROBERTSON, N. P. (2022). A role for the Epstein-Barr virus in multiple sclerosis aetiology?. *Journal of neurology*, 1–2. Advance online publication. <https://doi.org/10.1007/s00415-022-11177-w>
- KELLENBERGER E. (2004). The evolution of molecular biology. *EMBO reports*, 5(6), 546–549. <https://doi.org/10.1038/sj.embor.7400180>
- KLINEOVA, S., & LUBLIN, F. D. (2018). Clinical Course of Multiple Sclerosis. *Cold Spring Harbor perspectives in medicine*, 8(9), a028928. <https://doi.org/10.1101/cshperspect.a028928>
- KOKS, G., PFAFF, A. L., BUBB, V. J., QUINN, J. P., & KOKS, S. (2021). At the dawn of the transcriptomic medicine. *Experimental biology and medicine (Maywood, N.J.)*, 246(3), 286–292. <https://doi.org/10.1177/1535370220954788>
- LEE Y. H. (2015). Meta-analysis of genetic association studies. *Annals of laboratory medicine*, 35(3), 283–287. <https://doi.org/10.3343/alm.2015.35.3.283>
- LEE Y. H. (2018). An overview of meta-analysis for clinicians. *The Korean journal of internal medicine*, 33(2), 277–283. <https://doi.org/10.3904/kjim.2016.195>
- LOWE, R., SHIRLEY, N., BLEACKLEY, M., DOLAN, S., & SHAFEE, T. (2017). Transcriptomics technologies. *PLoS computational biology*, 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- MILO, R., & KAHANA, E. (2010). Multiple sclerosis: geoeidemiology, genetics and the environment. *Autoimmunity reviews*, 9(5), A387–A394. <https://doi.org/10.1016/j.autrev.2009.11.010>
- MOULTON V. R. (2018). Sex Hormones in Acquired Immunity and Autoimmune Disease. *Frontiers in immunology*, 9, 2279. <https://doi.org/10.3389/fimmu.2018.02279>
- MUÑOZ-CULLA, M., IRIZAR, H., SÁENZ-CUESTA, M., CASTILLO-TRIVIÑO, T., OSORIO-QUEREJETA, I., SEPÚLVEDA, L., LÓPEZ DE MUNAIN, A., OLASCOAGA, J., & OTAEGUI, D. (2016). SncRNA (microRNA & snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent. *Scientific reports*, 6, 20126. <https://doi.org/10.1038/srep20126>
- NEKRUTENKO, A., & TAYLOR, J. (2012). Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature reviews. Genetics*, 13(9), 667–672. <https://doi.org/10.1038/nrg3305>
- NELSON, D. L., & COX, M. M. (2017). *Lehninger principles of biochemistry* (7th ed.). W.H. Freeman.
- OLIVEROS, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. <https://bioinfogp.cnb.csic.es/tools/venny/index.html>

- PAGE, M. J., MCKENZIE, J. E., BOSSUYT, P. M., BOUTRON, I., HOFFMANN, T. C., MULROW, C. D., SHAMSEER, L., TETZLAFF, J. M., AKL, E. A., BRENNAN, S. E., CHOU, R., GLANVILLE, J., GRIMSHAW, J. M., HRÓBJARTSSON, A., LALU, M. M., LI, T., LODER, E. W., MAYO-WILSON, E., MCDONALD, S., MCGUINNESS, L. A., ... MOHER, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews*, 10(1), 89. <https://doi.org/10.1186/s13643-021-01626-4>
- R CORE TEAM (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- SANTIAGO-RODRIGUEZ, T. M., & HOLLISTER, E. B. (2021). Multi 'omic data integration: A review of concepts, considerations, and approaches. *Seminars in perinatology*, 45(6), 151456. <https://doi.org/10.1016/j.semperi.2021.151456>
- TARCA, A. L., ROMERO, R., & DRAGHICI, S. (2006). Analysis of microarray experiments of gene expression profiling. *American journal of obstetrics and gynecology*, 195(2), 373–388. <https://doi.org/10.1016/j.ajog.2006.07.001>
- THOMPSON, A. J., BARANZINI, S. E., GEURTS, J., HEMMER, B., & CICCARELLI, O. (2018). Multiple sclerosis. *Lancet (London, England)*, 391(10130), 1622–1636. [https://doi.org/10.1016/S0140-6736\(18\)30481-1](https://doi.org/10.1016/S0140-6736(18)30481-1)
- VEGETO, E., VILLA, A., DELLA TORRE, S., CRIPPA, V., RUSMINI, P., CRISTOFANI, R., GALBIATI, M., MAGGI, A., & POLETTI, A. (2020). The Role of Sex and Sex Hormones in Neurodegenerative Diseases. *Endocrine reviews*, 41(2), 273–319. <https://doi.org/10.1210/endrev/bnz005>
- VLACHOS, I. S., ZAGGANAS, K., PARASKEVOPOULOU, M. D., GEORGAKILAS, G., KARAGKOUNI, D., VERGOULIS, T., DALAMAGAS, T., & HATZIGEORGIOU, A. G. (2015). DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic acids research*, 43(W1), W460–W466. <https://doi.org/10.1093/nar/gkv403>
- VLACHOS, I. S., PARASKEVOPOULOU, M. D., KARAGKOUNI, D., GEORGAKILAS, G., VERGOULIS, T., KANELLOS, I., ANASTASOPOULOS, I. L., MANIOU, S., KARATHANOU, K., KALFAKAKOU, D., FEVGAS, A., DALAMAGAS, T., & HATZIGEORGIOU, A. G. (2015). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic acids research*, 43(Database issue), D153–D159. <https://doi.org/10.1093/nar/gku1215>
- VOSKUHL R. R. (2020). The effect of sex on multiple sclerosis risk and disease progression. *Multiple sclerosis (Houndmills, Basingstoke, England)*, 26(5), 554–560. <https://doi.org/10.1177/1352458519892491>
- VOSKUHL, R. R., SAWALHA, A. H., & ITOH, Y. (2018). Sex chromosome contributions to sex differences in multiple sclerosis susceptibility and progression. *Multiple sclerosis (Houndmills, Basingstoke, England)*, 24(1), 22–31. <https://doi.org/10.1177/1352458517737394>